DOCUMENT RESUME

ED 400 297 TM 025 614

AUTHOR Johanson, G.; And Others
TITLE The Evaluation of the Lead Teacher Project.
PUB DATE Apr 96
NOTE 5p.; Paper presented at the Annual Meeting of the American Educational Research Association (New York, NY, April 4-8, 1996).
PUB TYPE Reports - Evaluative/Feasibility (142) -- Speeches/Conference Papers (150)

EDRS PRICE MF01/PC01 Plus Postage.
DESCRIPTORS *Academic Achievement; Course Evaluation; Curriculum; Elementary Education; *Elementary School Teachers; *Inservice Teacher Education; Instructional Leadership; *Mathematics Instruction; Partnerships in Education; Program Evaluation; *Science Instruction; Standards; Student Attitudes; Teaching Methods; Workshops
IDENTIFIERS California Achievement Tests; Ohio University

ABSTRACT

The Lead Teacher Project at Ohio University included 84 teacher participants who represented 42 schools from 13 school districts in Appalachian Ohio. The project was designed to enhance the teaching and learning of science and mathematics at the elementary school level by helping the teacher participants develop skills, knowledge, attitudes, and leadership capabilities. The project, funded by federal, state, and local sources, used the resources of a range of experts at Ohio University in working partnerships for teacher training. The evaluation of the project was, in many ways, as diverse as the project itself. It was first necessary to evaluate the effectiveness of the teacher-education courses. This was done through a course evaluation questionnaire, weekly surveys, and informal discussions. Some analyses used the teacher as the unit of analysis and others used the teacher's students as the point of analysis. Student achievement was measured through the California Achievement Test, a process skills instrument created for the evaluation, and a curriculum standards survey for teachers was designed and implemented. Student attitudes toward mathematics and science were recorded by a project-developed instrument. Growth in teacher leadership skills was assessed through still another project-developed instrument. In general, the project was evaluated with respect to student performance, teacher performance, school building changes, and leadership. These measurements demonstrate the effectiveness of the project. (SLD)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)
☑ This document has been reproduced as received from the person or organization originating it.
☐ Minor changes have been made to improve reproduction quality.

• Points of view or opinions stated in this document do not necessarily represent official OERI position or policy.

PERMISSION TO REPRODUCE AND DISSEMINATE THIS MATERIAL HAS BEEN GRANTED BY
GEORGE E JOHANSON
TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

The Evaluation of the Lead Teacher Project

A paper presented at the annual meeting of the American Educational Research Association, New York, 1996

G. Johanson, R. Martin, C. Gips, B. Beach, & S. Green
College of Education, Ohio University

ED 400 297

The *Lead Teacher Project* at Ohio University included 84 teacher participants who represented 42 school buildings from 13 school districts in Appalachian Ohio. This cohort of teachers participated for the life of the project from 1990-94. Called Lead Teachers (LTs), the participants served more than 6,000 of their own students during this time and influenced 750 other teachers and 15,000 students from other classrooms. The project was designed to enhance the teaching and learning of science and mathematics at the elementary school level. This goal was achieved by helping the teacher participants to develop science and mathematics knowledge, skills, attitudes, and most importantly, leadership capabilities for their schools and fellow teachers. LTs were expected to disseminate their new knowledge, skills and attitudes to their peers and students. Each LT enrolled in either science or mathematics content and pedagogy, and they all studied and practiced leadership. LTs became school change agents in this urgent time of reform in mathematics and science. The *Lead Teacher Project* evaluation is worthy of reflection for the way in which it demonstrates the largely successful efforts of an integrated and collaborative evaluation of the teaching of science and mathematics at the local school and district level.

The project was funded at more than $1.5 million. Primary funding was received from the National Science Foundation, with supplementary funds coming from the Ohio Department of Education, Ohio Board of Regents, Ohio University, and from the consortium of participating schools. The project used the resources of a wide range of experts at Ohio University, such as faculty in engineering, meteorology, chemistry, physics, geology, botany, zoology, mathematics, and education. Additional staff were provided by the Ohio Department of Education, the participating schools, and other public agencies, e.g., Ohio Department of Natural Resources. All staff teamed to provide essential training for teachers in order to transform school curricula and teaching practices. The evaluation of this multifaceted approach to educational reform through working partnerships forms the basis of this paper in which we attempt to summarize:

1. the development of the project and an overview of the evaluation
2. instrument development and validation
3. the conclusions that were drawn from and about the on-going evaluation
4. and recommendations for similar efforts at evaluation.

Evaluation Overview

The evaluation of the *Lead Teacher Project* was, in many ways, as diverse as the project itself. Since there were significant instructional components for the teachers in the project, it was necessary to evaluate the effectiveness of these courses. This was largely done using a course evaluation questionnaire, weekly survey questionnaires, and through informal discussions with the participating teachers. These evaluation efforts were shared with the staff and were largely formative. The objectives of the program called for evaluating achievement in, and attitudes toward, science and mathematics. Teacher leadership also had to be evaluated. The decision was made to both pre- and post-test student groups for each of the three years of the project.

Originally, it was hoped that we could enlist the more limited participation of several surrounding schools and school districts that were not taking part in the program to serve as a comparison group. This proved to be impossible largely because of the lack of commonly used pre-existing measures and the extra burden of cost. It was decided to use those teachers (and their students) receiving the mathematics component of the project as controls for the teachers receiving the science component (and their students) and vice-versa. We realized that the use of such internal comparisons would make the interpretation of the evaluation data more difficult. In addition, for some portions of the evaluation, the student was the unit of analysis; for others, the classroom (or LT) was the unit of analysis.

TM025614

BEST COPY AVAILABLE

By collapsing the student data for each classroom, it was possible to compare student and LT measures. There were at least three drawbacks to using the LT as the unit of analysis. First, the number of subjects (and thus the power of the statistical tests) was drastically reduced; second, much information was lost collapsing across students within a class; third, different grades used somewhat developmentally different instruments or forms for most measures. However, the LT was the focal point of the project and many of the project goals. The analyses with the LT as the unit of analysis were therefore central to our reaching conclusions concerning these goals.

In discussing the need to evaluate student achievement, the project staff were initially torn between the necessity for a credible tool (the standardized test) and the desire to assess more of the higher level cognitive skills (process or problem-solving skills) that the project was dedicated to developing. We compromised by doing both. Gains on the relevant math and science sections of the California Achievement Test (CAT) were used to track the increases in the LT students' knowledge of science and mathematics. The costs involved were reduced by scoring the CATs in-house. The process skills assessment required the creation of an additional instrument.

Overall, the LTs perceived the process or problem-solving skills of the students as being closer to 'needs improvement' for each of the three Fall assessments and closer to 'satisfactory' for each of the Spring assessments. There was a significant interaction between the training (mathematics or science) and the three gain scores for math and science for each of the three years of the project. For all three years and in both disciplines, the gains were all positive. The highest process skill gains occurred in the first year of the project, but if broken down by grade level, the pattern is more complex.

A curriculum standards survey was designed with the help of the LTs to measure the degree of compliance with established recommendations regarding curriculum standards in mathematics and science. Curriculum standards self-reports were made annually by each LT for the duration of the project.

We might interpret the change in level of adherence to curriculum standards as going from a bit more often than half of the time to a near majority of the time over the three years of the project. The third-year gains were significant when analyses were conducted with all of the LTs and when only those LTs who had remained in one discipline for the entire duration of the project were utilized. It was expected that these third-year gains would be reflected somewhat equally by both the mathematics and science LTs. This was not the case. There was a statistically significant interaction between adherence to the curriculum standards and the training/discipline of the LT. In fact, only the mathematics teachers made significant gains in curriculum adherence going from near 'half of the time' in adherence to almost a 'majority of the time'. The science teachers' curriculum standards scores remained acceptable, but almost constant over the life of the project. Perhaps this was because of the lack of a single state-approved curriculum model in science (such a curriculum did exist in mathematics).

Student attitudes toward science and mathematics were measured by another project-developed instrument. This was a questionnaire with 15 mathematics and 15 science attitude items. Two forms were developed, one for grades 1-3 and one for grades 4-6. Since it was desirable to equate scores across all grade levels, 10 items on each form were common for mathematics and another 10 for science. A professional artist was commissioned to create the pictures and a staff member in science and another in mathematics were enlisted to devise items.

For mathematics attitude, the science LTs were to play the role of a control group for the mathematics LTs. The roles would be reversed for science attitude. Overall, math and science attitudes tended to decrease slightly from the second to the third year of the project. For both attitudes, the students of the science teachers appeared to suffer the greater losses. A closer inspection of the data indicated that there were a few exceptional scores (relatively exceptional in the sense of very large or small standard scores) that accounted for virtually all of the differences between the attitudes of the students of the mathematics and science trained LTs. When these outliers were removed, there remained little or no difference between either years 2 and 3 or between math and science teachers. That is, mathematics and science attitudes were reasonably unchanged over the course of either the second or third year of the project and did not differ between the experimental and comparison groups.

Growth in teachers' leadership skills was measured on a competency based (58-item) instrument, also developed by the project staff, and completed once each year by both the LTs and their respective school principals. Since the teachers and principals would be involved for the entire three years of the project (as opposed to students who were with a LT for only one year), we did not feel the need for more than three leadership measures. We would come to discover that leadership was, at least for us, a somewhat more complex (or elusive) construct to evaluate with a survey instrument.

In all three years of the project, the principal's rating of LT leadership was slightly above the self-rating of the LT. The final year of the project shows a significant increase in mean leadership score over both the first and second years for teachers, while the second year increase over the first is not significant. The Year 2 to Year 3 gain for the principal's rating only nears significance. A repeated measures analysis yielded significance only for the teacher gains.

In addition to these more formal and largely quantitative data, detailed documentation of virtually all activities through observation, professional achievement surveys, and formal and informal interviews and discussions at gatherings provided a picture of the tasks taken on by the LTs and their accomplishments. These somewhat less formal data were especially relevant to our evaluation of building changes. That is, one of the larger project objectives related to the dissemination of teacher skills and instructional/curricular information throughout the building (and possibly even further) making use of the leadership skills of the LT.

Instrument Development and Validation

A survey instrument was developed with the help of the participating teachers and project staff to (first define and then) assess the process skills of the elementary students. This was used twice annually (Fall and Spring) to collect teacher judgments of the process skill attainment of each of their students. These process skills were a primary point of the project and one of the main uses of the standardized test (CAT) scores was to validate our use of teacher judgment. We were not certain that the participating teachers would be able to give unbiased estimates of these skills; we were also concerned with the credibility of our efforts in the area of process skills. The use of a well-respected standardized test was seen to be an asset in both regards. Two forms of the CAT (E and F) were available for grades 3-6. In those classes, one form would be used in the Fall and the other in the Spring. A linear equating study was conducted each year and the scores were brought to the same scale prior to computations of gain. Item analyses and reliabilities for all CATs and process skills forms were satisfactory. Correlations between subscores (math computation, math application, and science) on the CAT and the related process skills scores were consistently large providing evidence for the validity of our process skill measures.

A curriculum standards survey was also developed with the help of the participants in the project. Since both the process skills and the curriculum standards varied with both curriculum (math or science) and grade level (1-6), there was a need for a multitude of forms for both instruments. An concerted effort at consensus was made for the inclusion of items on these forms. Item analyses and internal consistency reliabilities were again satisfactory. Correlations with achievement and attitude measures were mixed.

The four surveys of student attitudes towards science and mathematics (two grade levels) had good internal consistency reliabilities and no flawed items. Surprisingly, the factor structure of all four forms was virtually perfect with all of the math items loading on one factor and all of the science items loading on the only other factor. Construct (or factor) validity was evident.

While the survey of leadership skills seemed straight forward, there were problems. Correlations between principal and teacher scores were virtually zero. The leadership scores did not relate well to any of the achievement measures (in fact, some of the correlations were decidedly negative). With 58-items and typically fewer than 80 teachers, factor structure could not be investigated. There was evidence that some principals were reluctant to evaluate teachers poorly. Internal consistency reliabilities were satisfactory.

Conclusions and Recommendations

In general, project objectives could be classified as those relating to: pupil performance, LT performance, building changes, or leadership. With pupil performance, we had gains in CAT scores and gains in the process skills. The student was the unit of analysis and we had to do these analyses annually since students were with a particular LT for only one year. Positive gains and positive relationships were observed for each of the three years of the project.

The measures of LT performance were largely (but not exclusively) the annual measures of adherence to curricular standards and the leadership measures from the teacher and his or her principal. The unit of analysis here is the classroom or teacher. To relate these measures to student achievement, the student data were aggregated by classroom. Relationships with these other measures were generally positive but rather weak. The complications in the evaluation process (anticipated and otherwise) seemed to be due to a rather wide variety of causes.

BEST COPY AVAILABLE

1. The need to use 'bubble' sheets for data collection when many of the students were too young to complete these (the teachers in the lower grades recorded all observations on these sheets) caused an imbalance of effort on the part of participating teachers. Better anticipation of some of these inequities could have resulted in differentiated rewards.
2. The delivery of materials in a timely manner (and to the correct person) in a rural area and the prompt return of such materials was first done personally and later by mail. The mail was less costly and procedures were already in place in building for handling mail.
3. Having teachers reassigned to other classes, duties, or grades during the three years of the project did cause missing data for life-of-the-project analyses. Having a small number of replacements available might have been possible, had we anticipated this attrition.
4. A continued commitment to the evaluation required much time and effort on the part of the LT. As expected, some did not rise to the challenge as well as others.
5. The method for selection of LTs was not perfectly adhered to by all of the school districts. The project staff could have had more direct participation in the selections.
6. Using internal controls when one of the objectives of the program was dissemination and transfer was unwise. Attitude measures may have been most affected. In hindsight, an improvement might have been had with at least a small external comparison group.
7. Surveys do not always provide the best or most complete data for all problems. The data on leadership was especially poor using surveys.
8. Having students in classes with LTs for only one year and evaluating gain scores is a rather severe test of program effectiveness. Multiple years could perhaps have been arranged for at least a small contingent of the students.

On the positive side, there is also a variety of observations to be made.

1. We are reasonably convinced that process or problem-solving skills can be accurately assessed by classroom teachers using judgmental methods. The use of a standardized test, however, gave credibility to these measures. There are considerable savings to be had by using teacher judgment as opposed to the alternative of performance measures.
2. The use of the CATs did not seem to cause the participants to focus on material less related to the project objectives. Local scoring was successful but, of course, did not permit comparisons with national norms.
3. There were numerous meetings and other opportunities for formative feedback with respect to program objectives. Participants were encouraged to contact the evaluation team when there were questions and/or problems. This seemed to have been successful.
4. Having some tangible rewards in place for the participants encouraged participation in all aspects of the project and project evaluation.
5. While the use of 'bubble' sheets did present some initial difficulties, we are convinced that there was a substantial net savings in terms of time and effort.
6. The creation of so many of the instruments was a mixed blessing. The focus was where we desired (content validity was high), but much time and effort was required for the validation process. In the end, the commitment of the LTs to measures that they had helped construct made the added effort seem worthwhile.

By and large, LTs' students demonstrated an increase in their mathematics and science knowledge and skills on the California Achievement Test over several measurements during the project's lifetime. Student process skills in science and mathematics showed significant growth as well and these correlated positively with the CAT gains. Adherence to desirable curriculum standards was satisfactory. LTs increased their leadership skills in their own eyes and sometimes in the eyes of their school principals who evaluated them on the same instrument. Documentation of leadership activities shows that participants in the program completed more than 600 staff development programs that benefited their peers and more than 300 special events for students in mathematics or science. At last count, these same teachers attracted more than $170,000 in additional funding through their own initiatives for their school districts (a 54% proposal funding rate) and much of this effort continues today.